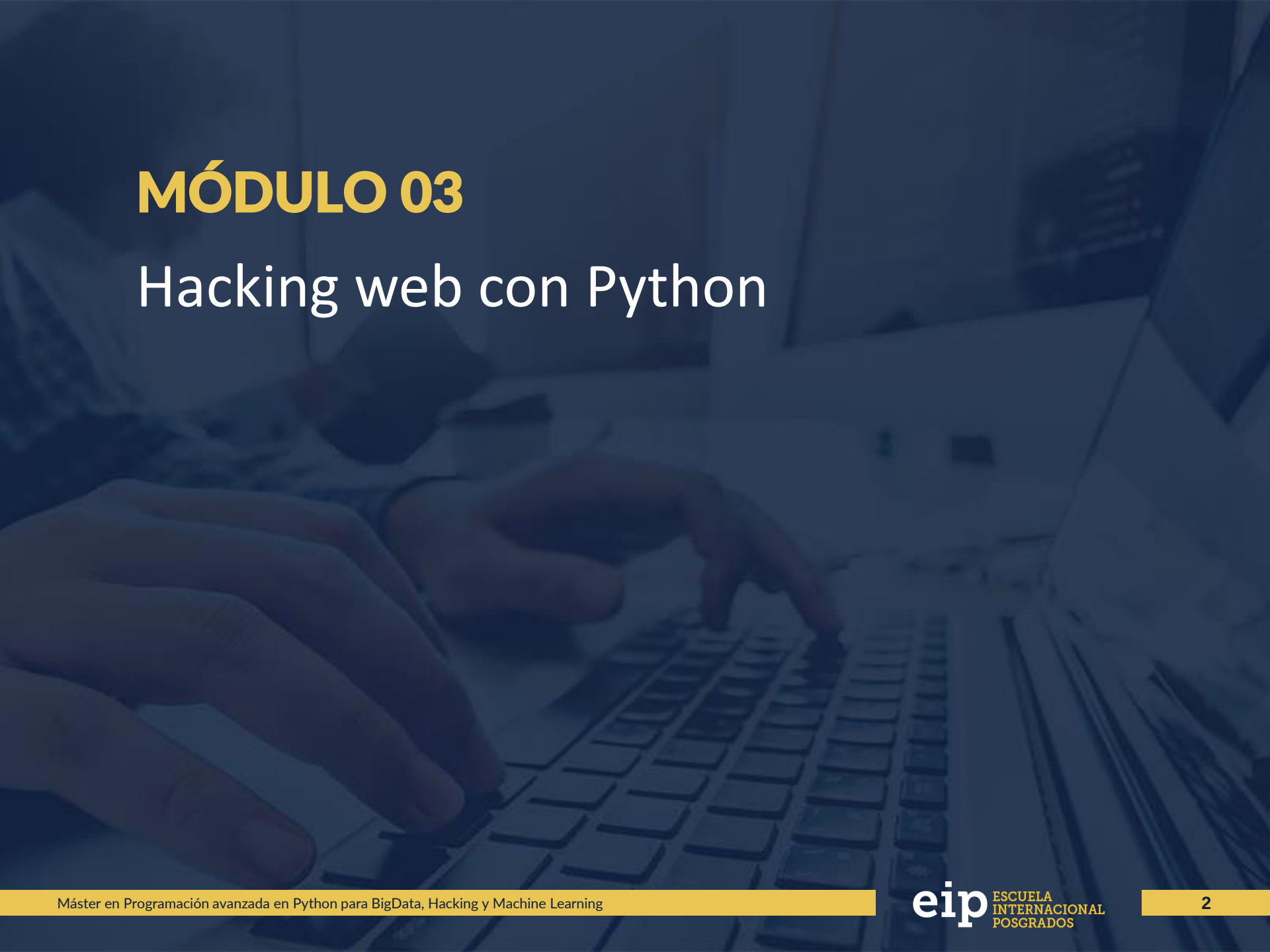
ESCUELA INTERNACIONAL POSGRADOS

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Hacking y Pentesting con Python



MÓDULO 03

Hacking web con Python

Clientes y servicios web con Python

El protocolo HTTP (transferencia de datos “hyper-text”) define las reglas que deben de seguir clientes, proxies y servidores para el intercambio de información.

La especificación más reciente (RFC) de este protocolo hasta la fecha, es la versión 2.0/3.0

Se trata de un protocolo sencillo, donde los clientes realizan peticiones y los servidores emiten respuestas.

Para almacenar información relativa a una transacción HTTP, se pueden utilizar cookies (valores almacenados en el lado del cliente) o sesiones (espacios de memoria temporal reservada para almacenar información sobre una o varias transacciones HTTP en el lado del servidor).

Clientes y servicios web con Python

Existen 8 métodos que pueden ser utilizados para acceder a los recursos de un servidor por parte de un cliente. Los métodos disponibles son: POST, GET, HEAD, PUT, DELETE, TRACE, OPTIONS, CONNECT.

El protocolo define un conjunto de códigos de respuesta que le permite a un cliente, conocer el resultado de la transacción HTTP iniciada con su petición. Típicamente los códigos HTTP se categorizan de la siguiente forma:

HTTP 100: Conexión Rechazada.

HTTP 2xx: Operación Exitosa.

HTTP 3xx: Redirección.

HTTP 4xx: Error por parte del cliente.

HTTP 5xx: Error por parte del servidor.

Gestión de contenidos web

La información que se transmite entre clientes y servidores web cuenta con un formato y unas reglas conocidas por ambas partes.

Dicha información normalmente se transmite en la forma de documentos XML, JSON o HTML.

Los principales contenidos web se encuentran disponibles en estructuras jerárquicas, de tal forma que se puedan manejar sin mayores problemas desde cualquier librería o lenguaje de programación.

Los documentos en formato XML y HTML han sido los más populares en el mundo web, sin embargo, tras la llegada de los servicios REST el formato JSON se encuentra muy extendido y es ampliamente utilizado en miles de aplicaciones web en Internet.

Gestión de contenidos web: BeautifulSoup.

Se trata de una librería que permite la gestión de documentos con estructuras fijas, como por ejemplo documentos XML, HTML o JSON y que se puede implementar un fácilmente en cualquier script con Python.

Depende de un parser adecuado para su correcto funcionamiento, el parser por defecto y el más recomendado es "lxml" el cual soporta documentos HTML, XML y JSON.

A la hora de instalar ésta librería mediante PIP, es importante tener en cuenta que se debe indicar "beautifulsoup4" como nombre del componente a instalar.

Gestión de contenidos web: Mechanize.

Librería que permite crear clientes web ricos en funcionalidades, permitiendo el envío de múltiples peticiones HTTP.

Tiene las clases y elementos necesarios para crear un navegador web simple que permita interactuar con sitios web.

Ideal para crear aplicaciones que permitan realizar pruebas de pentesting web de forma automatizada.

No soporta Javascript. Se puede utilizar junto con Selenium proveer de dicho soporte.

MUCHAS GRACIAS POR SU ATENCIÓN

eip ESCUELA INTERNACIONAL POSGRADOS

decheverri@grupomainjobs.com

LINKEDIN

Daniel Echeverri
https://www.linkedin.com/in/adastra1/

twitter.com/eiposgrados

facebook.com/eiposgrados

instagram.com/eiposgrados